Towards Automated and In-Situ, Near-Real Time 3-D Reconstruction of Coral Reef Environments

David M^{c}Kinnon, Hu He, Ben Upcroft and Ryan N. Smith
School of Engineering Systems, Queensland University of Technology, Brisbane, QLD 4000, Australia
Email: d.mckinnon,ben.upcroft,ryan.smith@qut.edu.au, hu.he@student.qut.edu.au

***Abstract*—Coral reefs are biologically complex ecosystems that support a wide variety of marine organisms. These are fragile communities under enormous threat from natural and human-based influences. Properly assessing and measuring the growth and health of reefs is essential to understanding impacts of ocean acidification, coastal urbanisation and global warming. In this paper, we present an innovative 3-D reconstruction technique based on visual imagery as a non-intrusive, repeatable, *in situ* method for estimating physical parameters, such as surface area and volume for efficient assessment of long-term variability. The reconstruction algorithms are presented, and benchmarked using an existing data set. We validate the technique underwater, utilising a commercial-off-the-shelf camera and a piece of staghorn coral, *Acropora cervicornis*. The resulting reconstruction is compared with a laser scan of the coral piece for assessment and validation. The comparison shows that 77% of the pixels in the reconstruction are within 0.3 mm of the ground truth laser scan. Reconstruction results from an unknown video camera are also presented as a segue to future applications of this research.**

I. Introduction

Coral reefs are among the most diverse and biologically complex ecosystems on earth, supporting approximately 33% of marine fish species. These fragile ecosystems form the foundation for a variety of marine organisms and are crucial to supporting human life as well. Reefs are underwater structures composed of calcium carbonate secreted by vast colonies of tiny living animals called corals. These complex ecosystems prosper in warm, shallow, clear, sunny and agitated waters, and have been referred to as *rainforests of the sea*, forming some of the most diverse ecosystems on Earth. Although they occupy less than 0.015% of the oceans, they contain more than 25% of all known marine species [1]. Coral reefs have a significant socio-economic footprint, with their annual value estimated at >USD 375 billion. However, these are fragile ecosystems, that are under enormous threat. Recent research has estimated that up to 70% of the world's reefs have been threatened or destroyed by a variety of stresses, including climate change, ocean acidification, overfishing, overuse of reef resources, urban and agricultural runoff and water pollution [2]. Properly assessing the growth or decline of reef environments can indicate the health of the surrounding ecosystem and is important in broadening our knowledge of the effects of global climate change.

Computing an accurate estimate of the total surface area, volume and mass of an organism is considered of fundamental and practical importance in benthic ecology, specifically for understanding the complex dynamics of energy flow, cycling of organic matter, and carbonate production in aquatic ecosystems. A specific example for such a necessity lies in the study and assessment of coral reef ecology. Here, correlation between biomass, symbiotic dinoflagellate density, chlorophyll concentration and respiration rates can be directly related to surface area and volumetric measurements. Although physical measurements seem easier to acquire than biological or chemical measurements, prior and existing techniques for estimating the physical parameters of coral colonies are problematic, time-intensive, can require complicated laboratory procedures and need repeated sampling. Few non-intrusive methods exist for *in situ* estimation, and many methods resort to destruction of the colony.

Fig. 1. Still frame image from a video sequence of a coral reef environment.

In this paper, we propose a non-intrusive, and repeatable *in situ* method for generating a 3-D reconstruction of a coral reef environment for the purpose of estimating physical parameters, such as surface area and volume for the assessment of long-term variability. This study is an initial investigation into the *in situ* utilisation of our algorithms for 3-D reconstruction of complex underwater environments. Future work will implement the proposed techniques onto Autonomous Underwater Vehicles for automating large-scale, persistent monitoring efforts of aquatic environments, and specifically coral reefs.

II. Background

Our long term goal is to develop hardware and algorithms that are deployable on an AUV that provide scientists with a synoptic view of the reef environment surrounding on-going detailed quadrat and photo-quadrat studies. Overall, the aim of the proposed method is to satisfy ALL of the following requirements:

(i) Works for images recorded in moderately turbid waters (visibility ~ 5 m), with non-uniform lighting, and with foreign objects (e.g., fish) that could interfere in the scene.
(ii) Allows large data-sets and the investigation and analysis of significant areas of reef
(iii) Easily gather data *in situ* with COTS equipment
(iv) Deploy the system with a human diver or on an AUV for automated collection
(v) Obtain measurement accuracies on the order of millimetres
(vi) Perform the entire process including object reconstruction, and possibly object classification, on-line and in near-real time.

With the exception of Item (vi) above, there are existing methods that can accomplish one or multiple of the outlined requirements. However, there is no existing technique that can satisfy requirements $(i - v)$, and be performed in near-real time.

A. In Situ Sampling Approaches

There is a vast research area devoted *only* to the subject of sampling coral reefs [3]–[5]. Many methods have been developed, implemented and updated over the course of multiple decades. However, this sampling question remains an active area of research today. Similar to other sampling problems, the selection of a specific methodology is intrinsically linked to the spatial and/or temporal scales of the feature of interest. For example, a different method would be employed for measuring fluctuations in living coral cover than that used for studying annual sea surface temperature variability in the Great Barrier Reef.

In this study, we are interested in developing a 3-D reconstruction technique for marine, biological targets for automated, large-scale mapping and monitoring applications. The survey region has a spatial scale on the order of 10s of km^2. Physical properties under study vary on a temporal scale on the order of months. In particular, a practical application is the quantitative computation of surface area and volume of coral structures, along with an estimation of coral cover. Existing methods for such quantitative sampling are primarily performed by human divers, and are time-consuming, with accurate and precise methods taking considerable amounts of time. Thus, such sampling is usually limited to multiple $\sim 2 - 5$ m^2 areas over a spatially limited area of the reef. Augmenting these exiting human-based sampling techniques with the deployment of autonomous robotic assets can fill a substantial gap in understanding of an entire reef ecosystem by providing data at unprecedented spatiotemporal resolutions and providing a synoptic view of the entire region of interest. Currently, there are four principal methods utilised in the field for *in situ* sampling of coral reefs; quadrats, photo-quadrats, line intercept transects, and underwater video transects.

Quadrats are extensively used for sampling in all branches of ecology. These framed regions usually have an area of 1 m^2 and are divided into a uniform grid of 100 segments (i.e., 100 10 cm x 10 cm cells). Data are acquired relatively rapidly and cheaply in the field, however quadrat sampling (i) cannot be used to measure spatial relief (rugosity), (ii) make if difficult to sample large branching corals, e.g., elkhorn coral *Acropora palmata*, and (iii) only provide data on a 2-D surface, thus underestimating coverage of features which have a predominant orientation in the vertical plane, e.g., soft corals.

Photo-quadrats are good for coral monitoring programs. They can provide accurate information on reef cover, especially if stereo imagery is collected, however analysis of photographs can be time consuming. Application of the proposed reconstructions methods and classification methods that are under development will significantly reduce the analysis time for this technique.

Line intercept transects are fairly rapid to deploy in the field. A fibreglass tape measure is laid close to the reef contour and the length (cover) of each reef category is recorded. The cover of each category is calculated by the ratio of number of points per category to the total number of points. The main limitation with line intercept transects is that it under-samples heterogeneous areas, e.g., areas of scattered corals.

Underwater video is well-suited to field survey because large areas can be covered quickly and the method can be used without extensive training. It also has the additional advantage of producing a permanent visual record of the data. A variety of methods exist for using video; the authors of [6] have used video for monitoring coral reefs of the Great Barrier Reef, and provide a detailed description of the method. Basically, the camera pointed directly at the seabed at an altitude of 1 m and 1.5 m. Divers are either towed or swim with the camera to gather data. Currently, the main drawbacks are the cost of deployment, equipment and processing facilities.

B. 3-D Coral Reconstruction Approaches

A variety of examples of 3-D reconstruction methods applied to coral reef environments can be found in the literature. The authors of [7], [8] demonstrate accurate reconstructions using x-ray tomography and visual imagery, respectively, however the methods cannot be performed *in situ* . The authors in [9] describe a laboratory method for calculating surface area and volume, with an associated technique for estimating 3-D surface area of field corals from photograph or video imagery gathered in the field. There are non-invasive methods presented for measuring the surface areas and growth of branching [10], [11] or foliate corals [12], and the surface area of non-branching massive corals [13], [14]. These methods are not designed for large-scale mapping and analysis and utilise archaic image processing methods for the reconstruction. The result is that processing cannot be done on-line. Additionally, [10] requires significant infrastructure to be taken into the field, and has a detailed image acquisition procedure.

The significant advantage of the proposed techniques over existing methods is the increased accuracy with significantly reduced processing time compared to any other existing method. Additionally, implementation is not based on the specific hardware, utilising simple and cost-effective COTS cameras for data collection. The processing algorithms allow for gathered data to be compared immediately to previous measurements, providing essential information regarding the data quality and scientific importance to scientists in the field. This can translate to a significant cost savings by guaranteeing that all necessary data are gathered while on

site, rather than returning to the lab to find out that there are significant holes or poor data quality post priori.

C. General Image-Based Reconstruction

The first phase of the process (structure-from-motion) a.k.a. Visual Odometry (VO) has a rich history in both the Computer Vision and Robotics literature. There have been a number of different groups that have presented successful approaches for VO. Some initial efforts were focused on the use of single uncalibrated moving cameras [15]. Issues such as projective drift in addition to the very difficult task of practical auto-calibration, led subsequent authors towards a focus on calibrated methods for visual motion estimation [16], [17].

Further approaches to VO focus on the use of stereo-camera rigs [16], [18]–[21]. These approaches have been successful in demonstrating that accurate VO can be obtained on a robotic platform at high frame-rates both with [19] and without the use of additional GPS and INS information. In [21], [22] it is reported that a purely vision based method is capable of estimating the total motion of a robotic platform within $1-2$ m over ~ 1 km trajectories without the use of loop closure detection.

In this presentation, we would like to point out that either single or multiple configurations of cameras will potentially be adequate for underwater VO. However, dependant on the physical constraints on the deployment vehicle a multiple configuration of (at least two) cameras will almost always produce more accurate VO estimates than would a single camera. This observation has been confirmed repeatedly within the literature.

The problem of solving for a 3-D reconstruction given a set of images can be likened to the task of inverse ray-tracing from the Computer Graphics literature - given a set of images, what was the 3-D model that generated these? In this sense, given the precise locations and orientations of each camera from the VO stage. It is now possible to calculate a 3-D representation of the scene based solely upon this information.

The topic of 3-D reconstruction from images is well established in the Computer Vision literature. There are numerous different approaches to the problem, all of which have as a common goal an initial stage of establishing (on a per image basis) the depth of the scene elements associated with the pixels from the images. It is important to understand that once the depth of a pixel location has been determined, given that the camera location is known, it is straightforward to back-project that point into the scene resulting in a 3-D point on the surface of an object.

There are two main threads within the literature for this task. The first are feature based methods, these methods segment a finite number of salient points (corners) within the images (i.e., a subset of the total number of pixels) that can be reliably extracted and matched into neighbouring images. Methods in this category proceed by repeatably matching features from one image into as many images as is possible subject to visibility constraints (such as occlusion) within the scene. Once a set of matches between features has been established, the 3-D location of the features can be triangulated. The result of this process is a sparse set of 3-D points corresponding to the segmented image features [23], [24].

A second category of methods are multi-view stereo approaches. In this case, the goal of the algorithm is to produce, for every pixel in each image, an estimate of its depth. This is of course a more onerous task than that of recovering a finite set of features, however in practice, assumptions about the local continuity of surfaces can be employed to greatly reduce the complexity of this task. The output from a multi-view stereo approach is similarly a set of 3-D points, however there will typically be a far greater number of points due to the dense nature of per-pixel depth estimates [25]–[27].

Feature based methods have the advantage that they produce 3-D points that are associated with more than one image, this helps to reduce the duplication of 3-D point estimates contributed by neighbouring images. The disadvantage is that due to the sparse nature of the salient feature selection process, the total number of 3-D measurements may be two orders of magnitude less than those available from a multi-view stereo approach. The main advantage of the multi-view stereo approach is the large number of 3-D measurements that it produces. This is also somewhat of a disadvantage, seeing as it becomes more difficult to process this large number of points into a meaningful 3-D reconstruction.

The second stage within the 3-D reconstruction process is the fusion of the set 3-D points into a 3-D polygon model of the scene. Again, there are two main approaches to this task. These can be broadly classified as implicit and parametric methods. At this point, it should be noted that for some applications (such as obstacle avoidance and collision detection) the fusion of 3-D points into a polygon model may not be necessary, the availability of 3-D point measurements may by themselves provide enough information for subsequent processing.

Implicit methods of 3-D reconstruction [23], [25]–[27] seek to fuse all 3-D measurement data into a volume of finite size encapsulating the total extent of the object/scene. The volume may then be processed to approximate a solution for the surface that best fits the observed data. Implicit methods are suitable for small objects or scenes or cases where the fidelity of the 3-D model is not of the highest importantance. Memory limitations prohibit the use of implicit methods on larger scenes even when octree approaches are utilised to store the information in a sparse format.

Parametric methods of 3-D Reconstruction [24], [28] attempt to calculate some form of polygonal model as a first step (i.e. parametric representation) then through subsequent processing deform this 3-D model so that it best fits the calculated set of 3-D points and/or the observed correspondence data within the images. The great advantage of parametric methods is that only the final 3-D surface needs to be stored, in this sense they are more broadly applicable to large and small scenes. The great disadvantage of parametric methods is the considerable difficulty in maintaining the correct topology of a mesh throughout its evolution, it is this fact that has severely limited the adoption of this type of approach.

III. Algorithm Outline

The method used for our 3-D reconstruction consists of two distinct phases. The first phase is an automatic calibration (a.k.a. structure-from-motion) procedure. This consists of feature extraction and tracking, followed by camera pose

estimation routines and bundle-adjustment to refine the solution. The lens calibration can either be known a priori or approximated using auto-calibration routines to calculate the true values of the focal length, distortion etc. of the camera. In this work, we assume that the lens calibration has been determined prior to the deployment of the imaging device.

The second phase of the algorithm calculates dense polygonal models of the scene contents using multi-view stereo followed by an implicit 3-D reconstruction of the stereo data. We choose to use a dense multi-view stereo algorithm as we have observed this to provide a superior quantity and fidelity of 3-D points for subsequent surface reconstruction. Additionally, we fuse the 3-D points from the stereo processing with an implicit surface reconstruction algorithm. For this application, due to the relatively small size of the objects we are modelling, the implicit algorithm is a good choice capable of capturing a high degree of the surface detail.

A. Visual Odometry

The VO system utilised in this paper is based upon [22]. The system presented in [22] was shown to recover the VO of a set of cameras with a high degree of accuracy over long trajectories. Here we sketch an extension of this algorithm to also also handle the case of a single moving camera (see Algorithm 1).

As a summary, Alg. 1 consists of feature extraction (using GPU accelerated SURF features [29]), followed by GPU accelerated matching of the features between timepoints. In the single camera case, to initialise the reconstruction at least two images are tracked and an essential matrix is determined allowing for an instantiation the 3-D structure $\mathbf{X}^j$ for the sequence. In the multi-camera case, the structure can be initialised at the first time-point by triangulating the 3-D structure between all cameras k at time-point i. Following this the pose of each camera $\mathbf{P}_{k,i}$ is esimated from previously tracked structure using the 3-point resectioning algorithm (any of the approaches presented in [30] are suitable for this purpose). Finally, the pose estimates at each time-point are refined using a sliding window version of Bundle-Adjustment [31].

The VO system runs at frame-rates of up to 15 Hz, the resulting odometry is highly accurate and has shown to be robust to jerky and rapid motions of the camera, see [22] for further details.

B. 3-D Reconstruction

The camera pose information recovered from the VO stage can now be used to reconstruct a 3-D model of the scene. The algorithm we present is in three parts, the first part consists of a novel multi-view stereo approach that we call the Semi-Local algorithm. This algorithm has proven to be capable of generating highly accurate sets of dense 3-D points for each image.

The second part is a depth consistency checking process, utilized to remove erroneous depth estimates from the multi-view stereo output. The depth consistency check is performed at each stage of the multi-view stereo refinement process. This allows for a continual enforcement of the occlusion and visibility constraints from adjacent images to be applied (similar in effect to [23]). To our knowledge, this is the first instance of simultaneous enforcement of such constraints within a multi-view stereo refinement process.

Algorithm 1: Single/Multiple-Camera Visual Odometry

Require: Calibrated multiple-camera image set $\mathbf{I}_{k,i}$
Ensure: Estimates of the VO $\mathbf{P}_{k,i}$ and scene structure $\mathbf{X}^j$
1: **for** all time-points in sequence $i = 0, 1, 2, ..., n$ **do**
2: Detect SURF features for all images $\mathbf{I}_{k,i}$ at time-point i
3: Match features from $\mathbf{I}_{0,i}$ to $\mathbf{I}_{0,i-1}$
4: **if** *Multiple-Cameras* **then** For all $k > 1$. Guided matching of features from $\mathbf{I}_{k-1,i}$ to $\mathbf{I}_{k,i}$
end
Initialise the coordinate system (see text)
6: For all k. Robust 3-point pose estimation of the of $\mathbf{P}_{k,i}$
7: For all k. Single/Multiple-camera bundle-adjustment of $\mathbf{P}_{k,i}$
8: (Optional) Perform sliding window single/multiple-camera bundle-adjustment on frames $i - d$ to i to refine the pose and structure estimates
9: Initialise any new structure $\mathbf{X}^j$ visible over at least two time-points
10: **end for**

The final part consists of an implicit surface reconstruction algorithm. This component of the algorithm is capable of merging the dense 3-D points from each image into a parametric polygon model of the object or scene. Some care is taken to reduce the very large number of 3-D points extracted from the multi-view stereo stage down to more manageable number suitable for surface fitting.

Before presenting the algorithm, we will first establish some of the notation and define the variables used throughout. The goal of the algorithm is to produce a depth-map $\mathbf{D}_b$ for the base image $\mathbf{I}_b$ using M comparison images and their associated camera projection matrices $\mathbf{Q}_q = [\mathbf{N}_q|\mathbf{n}_q]$ (for $q = 1, \ldots, M$). The comparison images are selected by finding the cameras within a close proximity and viewing angle from the base camera.

Assume the base image has a camera projection matrix $\mathbf{P}_b = [\mathbf{M}|\mathbf{m}]$ that projects a 3-D homogeneous point $\mathbf{X}$ into its 2-D homogeneous location in the base image $\mathbf{x}_b$ (i.e., $\mathbf{x}_b \sim \mathbf{P}_b\mathbf{X}$) and the camera is located at $\mathbf{C}$ where $\mathbf{P}_b\mathbf{C} = \mathbf{0}$. The reprojection of a point into a comparison image $\mathbf{x}_q$ from a point ($\mathbf{x}_b$) in the base image with a depth d is,

$$\mathbf{x}_q = d\mathbf{N}_q\mathbf{M}^{-1}\mathbf{x}_b + \mathbf{N}_q\mathbf{C}. \quad (1)$$

We assume that some sort of photo-consistency metric is used to evaluate the cost $\mathcal{R}_q^d \in [0, 1]$ between a patch of pixels $\mathcal{P}_b$ from a location in the base image $\mathbf{x}_b$ to the corresponding patch $\mathcal{Q}_q^d$ in a comparison image at a given depth d. This patch is determined by a direct application of Eq. (1). The photo-consistency measure is normalised so that a value $\mathcal{R}_q^d = 0$ indicates similarity and 1 dissimilarity. In this paper, we use the Normalised Cross-Correlation (NCC) measure exclusively (with a window size $[3 \times 3]$) although there are numerous other alternatives to this that could also be used [32].

1) Semi-Local Multi-View Stereo: Our algorithm operates over an image-pyramid of k_l levels. To simplify calculations each subsequent level of the image-pyramid (k) is a power of 2 reduction from the previous level. We choose the first level

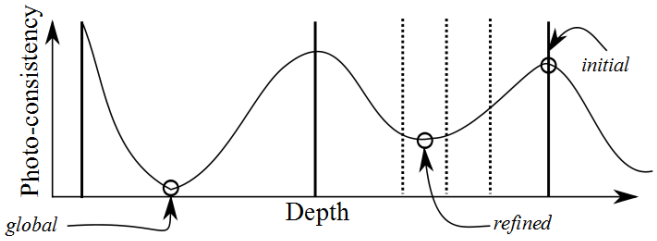


Fig. 2. Example of the photo-consistency cost function for a pixel. The locations of the discrete depth evaluations are indicated (bold vertical lines), as well as the global minima, refined (parabolic) update and the initial (discrete) depth evaluation.

to be the one resulting in an image size no smaller than 128 pixels in either dimension and increase the resolution from this point to the desired extent.

Note we do not downsample the original images in this process, instead we only attempt to reconstruct a sparse set depth measurements (for pixels in the base image) at each level. This fact is extremely important to maintain a consistent application of photo-consistency measure throughout each level of the refinement process.

The first stage in the hierarchical depth estimation is determining a suitable initial depth-map $\mathbf{D}_b^0$ for the smallest level of the image pyramid. Recall that the first level is many times smaller than the original image and consequentially the depth-map corresponding to this can be calculated very rapidly. For this purpose we have experimented with global methods for depth-map initialisation, principally the Semi-Global Matching (SGM) method [33], as well as simple Winner Take All (WTA) plane-sweeping approaches. We found that the WTA strategy produced sufficiently accurate initial estimates without the need to introduce the additional complexity (and superior accuracy) of SGM.

The initial depth-map from the first level can now be used as a guide in determing the depth maps for images of increasing pixel and depth resolution higher in the image pyramid. The approach used in this work follows from the algorithm presented in [34]. In [34] the assumption of local surface continuity is enforced by an iterative neighbourhood depth comparison/swapping procedure. This procedure sought to test and potentially replace depth estimates for a given pixel with a value from a neighbouring pixel with an improved photo-consistency score.

We extend this approach of testing neighbouring pixels to also evaluate a range of near by depths. This allows for an effective increase in the fidelity of the resulting depth estimates. The idea is that close-by pixels will likely share similar depths so these should be a good starting point for a more detailed search for the true depth of a pixel. Furthermore, due to the erratic nature of the photo-consistency function (over the depths of a pixel) the chances of a discrete plane-sweep procedure applied at the base level of the pyramid locating the true depth of a pixel are small. For this reason, we search depths at a finer increment (to that of the plane-sweep) to better ascertain the local nature of the photo-consistency function around a given depth. This concept is illustrated in Fig. 2.

At each iteration (n) a select group of candidate depths ($\mathcal{C}_p^{\cdot}$) are sampled from the surrounding pixels, note that the candidate depth set does not include the current depth estimate for the pixel. The patterns used to isolate the candidate depths are the same as those given in [34]. The implementation alternates between these candidate depth patterns at each iteration. For each of the candidate depths the surrounding depths $\mathcal{C}_p^+$ and $\mathcal{C}_p^-$ (generated by adding and subtracting the depth increment $\frac{1}{d_{\text{inc}}^k}$ from the original value $\mathcal{C}_p^{\cdot}$) are evaluated and the corresponding photo-consistency scores $\mathcal{R}_p^+$, $\mathcal{R}_p^{\cdot}$ and $\mathcal{R}_p^-$ determined.

If the smallest of the resulting photo-consistency scores is $\mathcal{R}_p^{\cdot}$ then a parabolic interpolation using the neighbouring scores and depth values is used to more accurately ascertain the corresponding depth minima. If either of the depths ($\mathcal{R}_p^+$ or $\mathcal{R}_p^-$) corresponds to the minima then they are simply selected as the updated value for the pixel. This iterative process is repeated n_l times on each level of the pyramid. We have observed experientially that this process of searching neighbouring pixel's depths for improved estimates results in both the ability of the algorithm to jump-out of local minima of the photo-consistency score, as well as dramatic increases in the depth fidelity for the pixels (via the refinement).

The value for d_{inc}^k can potentially be set at each pyramid level to govern the step size to be considered for each new depth candidate test. Greater values of the d_{inc}^k will allow the algorithm to potentially move faster to a solution at the risk of over shooting it's true depth value. Smaller values ensure that search proceeds more cautiously, but also increases the chance that a local minimum for the pixel's depth is found. It has also been noted experimentally that larger values of d_{inc}^k perform better on more oblique surfaces where the depth gradient is high so there may be some scope for this parameter to be data driven.

2) Depth Consistency Checking: Other authors have shown that depth-consistency checking can be used to great effect in multi-view stereo pipelines to reduce the incidences of bad depth estimates, and also as a means to regularise good depth estimates (filtering estimates from neighbouring views). It has previously been noted that regularising during the calculation of depth has a negative effect on the resulting depth-map accuracy [25]. Instead of attempting to alter the value of a depth estimate we employ a process of consistency checking strictly to remove the instance of spurious depth estimates as well as those that are not well supported by the surrounding views, similarly to the confidence-based fusion approach of Merrell et al. [35].

Previous authors have employed depth-consistency checking as a type of post-process to clean and enhance depth-maps upon completion of the stereo process. Here we argue that the neighbouring view's depth-maps provide important information that can be used during the depth refinement process to eliminate spurious and unsubstantiated depth estimates propagating through to the final depth-map. To this effect, the depth-consistency checking is inserted as a final stage at each level of the depth-map refinement. For this to be achieved, all neighbouring view's depth-maps must be refined simultaneously and then compared against each other at each level.

3) Implicit 3-D Reconstruction: Once a depth map has been extracted for each image, the next stage in forming a 3-D model requires that the individual depth maps be merged into a consistent polygon model. For this purpose, we introduce an algorithm focused primarily upon the reduction

Algorithm 2: Iterative Depth-Map Refinement

Input : Given a base image $\mathbf{I}_b$, M accompanying images $\mathbf{Q}_q$ and their camera matrices $\mathbf{P}_b$ and $\mathbf{P}_i$.

Output: Refined estimate of $\mathbf{D}_b^0$.

for $k = 0$ **to** $k_1 - 1$ **do**
  for $n = 1$ **to** n_1 **do**
    foreach *Pixel* $\mathbf{x}_b$ *of* $\mathbf{D}_b^{k,n}$ **do**
      Sample the patch $\mathcal{P}_b$ at this position;
      Sample a set of surrounding depth candidates $\mathcal{C}_p$;
      foreach *Depth candidate* $\mathcal{C}_p$ **do**
        for $q = 1$ **to** M **do**
          Sample the patches $\mathcal{Q}_q^+$, $\mathcal{Q}_q^\cdot$ and $\mathcal{Q}_q^-$;
          Accumulate scores $\mathcal{R}_p^+$, $\mathcal{R}_p^\cdot$ and $\mathcal{R}_p^-$;
        end
        if $\mathcal{R}_p^\cdot < \mathcal{R}_p^+$ *and* $\mathcal{R}_p^\cdot < \mathcal{R}_p^-$ **then** $\bar{\mathcal{C}}_p = interp(\mathcal{C}_p^+, \mathcal{C}_p^\cdot, \mathcal{C}_p^-)$;
        else if $\mathcal{R}_p^+ < \mathcal{R}_p^\cdot$ *and* $\mathcal{R}_p^+ < \mathcal{R}_p^-$ **then** $\bar{\mathcal{C}}_p = \mathcal{C}_p^+$;
        else $\bar{\mathcal{C}}_p = \mathcal{C}_p^-$;
      end
      Update $\mathbf{D}_b^{k,n}$ with the value of $\bar{\mathcal{C}}_p$ corresponding to the lowest score;
    end
  end
  Perform consistency-checking on $\mathbf{D}_b^{k,n}$ by comparing to neighbouring views $\mathbf{D}_q^{k,n}$
end

of the large number of 3-D points produced (from the multi-view stereo stage) down to a more manageable set. This smaller set of 3-D points still adequately captures the detail of the scene however the computation of the resulting implicit model is less onerous.

The first stage in reducing the size of the point set is extracting the 3-D points from the images. Due to the consistency checking and iterative refinement of the Semi-Local algorithm we use the fact that the depth-maps have a well enforced local surface consistency and perform a RANSAC-type [36] plane fitting procedure on the depth-maps themselves resulting in a set of oriented 3-D points for each image.

The oriented point sets from each image are then progressively added to an octree grid (of level L). Using a method closely related to [37], 3-D points that are in close proximity to each other are partitioned into progressively smaller sets of points (leaf nodes). Upon completion of this process, the partitions hold the accumulated mean and normal of all the contributing oriented points for the node. This typically results in a 10 to 20 times reduction in the size of the point set, whilst not significantly impacting the information content of the data.

The reduced oriented point set can now be fused into a 3-D polygonal model. For this purpose, we use the very popular Poisson Reconstruction [38]. This algorithm produces very accurate polygonal models from oriented 3-D points sets and has become the state-of-the-art method used for image-based 3-D reconstruction and laser scanner based point acquisition systems.

The Semi-Local stereo and consistency checking algorithms were implemented on the GPU and are capable of providing highly accurate depth-maps at better than 10 Hz depending on the size and number of images, desired depth accuracy and computation resources. The reduction of the corresponding set of points and Poisson Surface fitting portion of the algorithm are an offline process and can take up to 1 minute depending upon the octree level L, number of images and oriented points generated.

IV. Multi-View Dataset

To enable a quantitative evaluation of our multi-view stereo reconstruction algorithm, we collected a calibrated multi-view image set of a piece of Staghorn Coral, *Acropora cervicornis* in a shallow pool (Fig. 3) and a corresponding reference 3-D model from laser scan data.

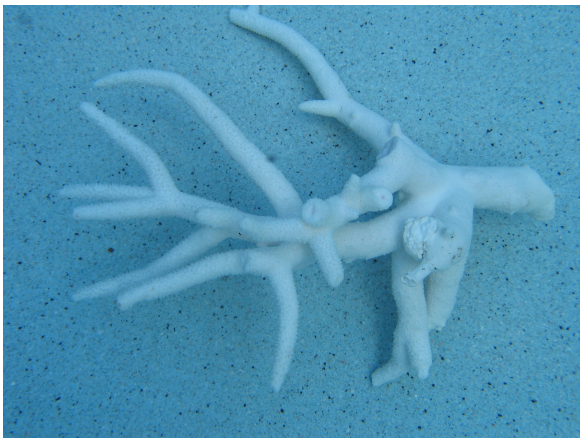


Fig. 3. One of the images of the piece of Staghorn coral from the collected underwater dataset.

Several other high quality multi-view datasets with registered ground truth models for terrestrial objects and scenes are publicly available [39], [40], but there is no ground-truth dataset for underwater analysis. The coral used in this study is very challenging for typical multi-view stereo reconstruction algorithms due to the non-concave structure and occlusions presented by the object. Furthermore, water characteristics can further degrade acquired images resulting in inaccurate reconstructions.

Images were captured under the water using a CCD camera with a resolution of 2272×1704 pixels at a distance of $\sim 1-1.5$ m. At this resolution, a pixel in the image spans roughly 3.144 μm on the surface of the coral which is approximately 35 mm$\times$25 mm$\times$15 mm in size.

A. Calibration

The camera was calibrated by imaging a planar calibration grid under the water from 21 viewpoints over a hemisphere using the Matlab Calibration Toolbox [41] to compute intrinsic and distortion parameters. A perspective camera model with radial distortion was assumed with intrinsic parameters $f_x, s, f_y, c_x, c_y, k_1, k_2$ where f_x, f_y is focal length along x and y axis, respectively, c_x, c_y are the location of principle point, s is the skew of the CCD array, and k_1, k_2 are the first two radial distortion parameters.

The settings on the camera were retained to ensure that the calibrated parameters remained the same for all successive datasets acquired for this paper.

The camera was approximately 0.8 m from the calibration grid for most viewpoints. Each square in the calibration grid is 70 mm×70 mm. The results for the calibration are:

$$\begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3083.39 & 0.00 & 1109.33 \\ 0.00 & 3089.55 & 806.67 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$$

$$\begin{bmatrix} k_1 & k_2 \end{bmatrix} = \begin{bmatrix} 0.05196 & 0.67905 \end{bmatrix}$$

B. Laser Reference Model

The 3-D reference model was captured using a Cyberware Model 3030/sRGB laser stripe scanner. The resolution of each scan was 350 μm. The laser scanner rotates completely around the object to provide a 3-D model. However, due to occlusions, multiple scans were acquired with the coral in different poses for each of the scans. Each 3-D laser model was aligned using Iterative Closest Point (ICP) [42] and then merged to form a single reference model (Fig. 4). Note that merging multiple scans increases the effective resolution due to an increased density of point measurements.

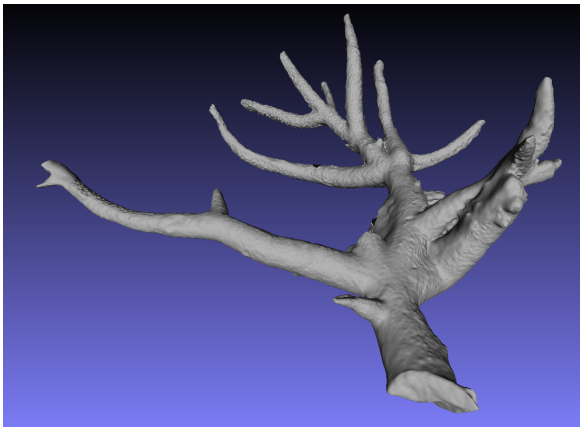


Fig. 4. Laser scan reference model of the piece of Staghorn coral.

C. Multi-View Image Dataset

Images of the coral were acquired without the calibration board (Fig. 3) as a test dataset for comparison to the 3-D laser reference model. We attempted to take images of the coral in a spiral pattern to form a comprehensive dataset with viewpoints from all angles. The images are sequential to enable the use of small baseline camera pose estimation algorithms.

Accurate multi-view 3-D reconstruction relies on the computation of the camera position and attitude (pose) at each viewpoint. The more accurate the pose, the greater chance of producing an accurate 3-D reconstruction. An estimate of the camera positions using the algorithms from Section III is shown in Fig 5.

V. Algorithm Validation

To evaluate the effectiveness of the presented scheme we performed a series of experiments with a publicly available ground-truth terrestrial dataset in addition to the multi-view aquatic dataset described in the previous section. The experiments were designed to determine the performance of the

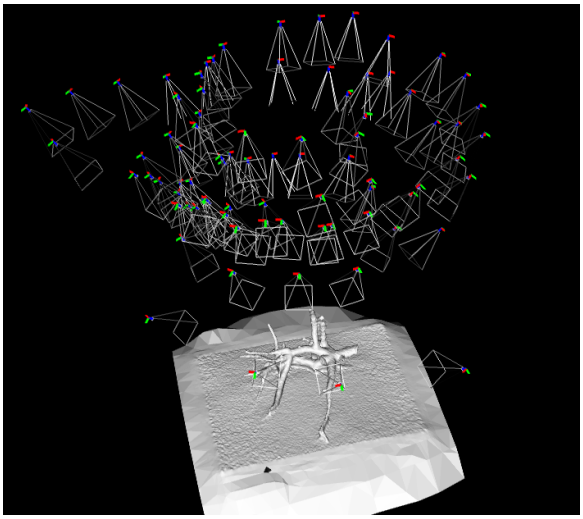


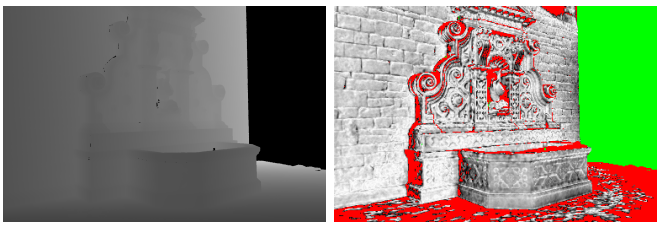
Fig. 5. Camera positions for the multi-view, underwater image dataset of the Staghorn coral.

presented algorithm over a fixed range of parameter values and settings. The parameters used were $M = 1$, $d_{\text{inc}} = 4096$, $k = 3$, $n_l = 3$, and $L = 10$.

A. Terrestrial

The ground-truth dataset is the *fountain* dataset from [43]. This dataset provides access to both the calibrated camera locations and images (11 images of resolution 3072×2048), in addition to LIDAR data (from a Zoller+Fröhlich IMAGER 5003) aligned to the same coordinate frame. The experiments use the ground-truth LIDAR data to create the corresponding ground-truth depth-maps, which in turn are used to evaluate the algorithm.

For each generated depth map, we gather statistics for pixels where the depth-refinement process and the LIDAR provide a depth estimate. Generally the proportion of pixels that result in a depth estimate is quite high ($> 90\%$), with the discarded pixels often occluded in the comparison image or associated with homogeneous regions where reliable depth determination is not possible using a small correlation window. The errors in the depth estimates are then computed and expressed as a multiple of the variance in the measurements. In this case, the variance $\sigma = 0.3$ mm. An example of the ground-truth depth-map and associated error image for the *fountain* dataset is shown in Fig. 6.



(a) Example image of the rendered ground-truth depth-map for the *fountain* sequence.

(b) Gathered error statistics: green denotes no ground-truth measure, red a gross error ($> 10\sigma$), white no error increasing towards black (10σ).

Fig. 6. Ground-truth depth map and error statistics for the fountain dataset.

The cumulative error of the depth estimates are illustrated as a percentage of total pixels in Fig. 7. This was compared

to the results from three state of the art techniques; a Winner Takes All (WTA) plane-sweeping approach, a Semi-Global Matching (SGM) method [33] and Cornelis' method [34]. The figure can be interpreted as $\sim 58\%$ of pixels from our method are within $3\sigma = 3$ mm of the ground truth depth. The results show that our approach achieves extremely accurate depth maps when compared to the other methods.

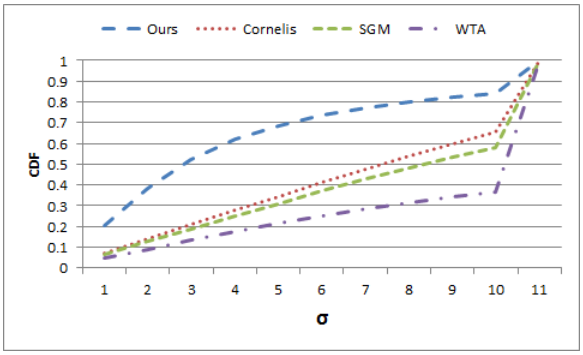


Fig. 7. The cumulative error (CDF) of the depth estimate as a percentage of total pixels. Pixels with an error greater than 10σ were combined into a single bin at 11σ.

B. Aquatic

The accuracy of the depth maps for the coral dataset were compared to the data from the 3-D laser reference model. An example of the ground-truth depth-map and associated error image for the coral dataset is shown in Fig. 8, where in this case the variance is $\sigma = 0.1$ mm.

The cumulative error of the depth estimates are also illustrated as a percentage of total pixels in Fig. 9. This figure can be interpreted as $\sim 77\%$ of pixels are within $3\sigma = 0.3$ mm of the ground-truth depth.

Similar accuracies have been achieved in the Seitz evaluation terrestrial dataset [39]. This dataset allows for a reasonable comparison, as the object and imaging distance are of similar magnitudes. The improved accuracies presented here can be attributed to the higher resolution images utilised for the reconstruction. We remark that the accuracy presented in the fountain dataset is somewhat less due to the increased imaging distance. Any *in situ* imaging of coral should attempt to maintain an imaging distance of ~ 1 m to achieve similar accuracies as presented here.

With accurate depth-maps, a high quality 3-D reconstruction can be obtained (Fig. 10) as described in Section III-B. The multi-view and laser models were subsequently aligned using Iterative Closest Point (ICP) [42] to evaluate the accuracy of the resultant multi-view model. The alignment parameters consisted of a rotation, translation, and uniform scale (required for monocular camera multi-view reconstructions). The average alignment error between vertices of the two models was 0.7 mm further indicating the high accuracy of our 3-D reconstructions.

(a) Example image of the rendered ground-truth depth-map for the coral sequence.

(b) Gathered error statistics: green denotes no ground-truth measure, red a gross error ($> 10\sigma$), white no error increasing towards black (10σ).

Fig. 8. Ground-truth depth map and error statistics for the underwater, coral dataset.

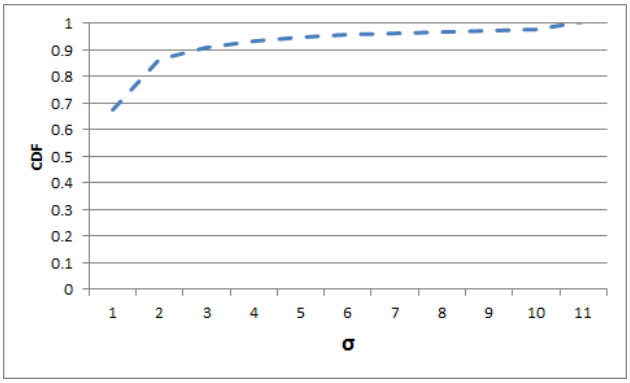


Fig. 9. The cumulative error (CDF) of the depth estimate as a percentage of total pixels. Pixels with an error greater than 10σ were combined into a single bin at 11σ.

VI. Conclusion

The majority of this paper presented the details of a set of innovative, multi-view stereo algorithms that provide rapid and accurate depth-maps for image-based 3-D reconstruction. The utility of these algorithms was demonstrated via their application on well-known, ground-truth, terrestrial datasets. We extend these results to applications in the underwater domain, specifically towards generating *in situ*, 3-D reconstructions of coral reef environments. Through the presented results, it was verified that the presented image-based reconstruction techniques are applicable in aquatic environments. This study is an initial investigation, utilising a piece of Staghorn coral (*Acropora cervicornis*) to test and validate the techniques *in situ*. Results from our image based 3-D reconstruction are compared to a laser scan, showing that

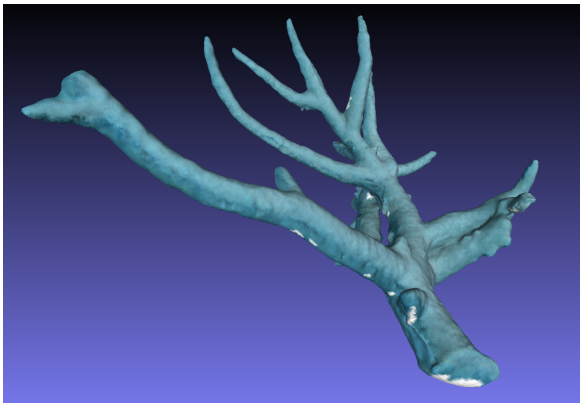


Fig. 10. Final 3-D reconstruction of the piece of Staghorn coral from the multi-view, underwater image dataset.

$> 90\%$ of the surface area of the coral is within 1 mm of the ground-truth. These results confirm that the methodology and approach are applicable for generating 3-D reconstructions of coral reef environments *in situ* for the purpose of estimating surface area and volume of reef constituents.

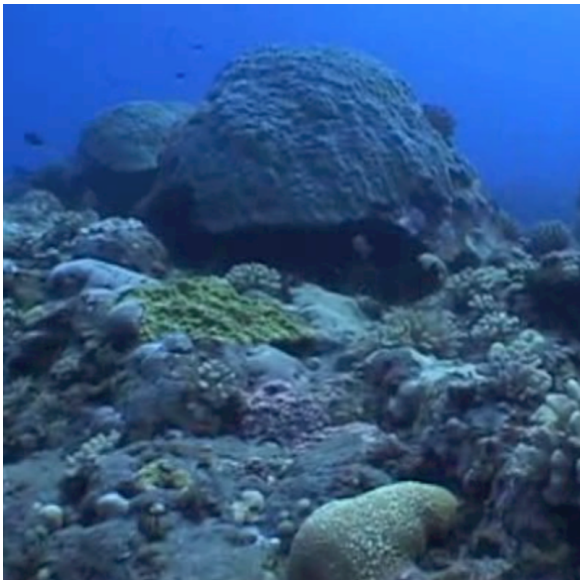
This method has broad applications for studying coral reef ecology and assessing long-term variability and change over large spatial extents, and with reduced manual effort in image processing and data analysis. Based on the accuracy and precision required for a desired science application, adjustments to the processing techniques presented may need to be adjusted, e.g., number of images taken, resolution of mesh utilised to build the reconstruction, etc.

The underwater coral dataset, along with ground-truth laser scan are freely available, and can be acquired by contacting either Ryan N. Smith (ryan.smith@qut.edu.au) or Ben Upcroft (ben.upcroft@qut.edu.au).

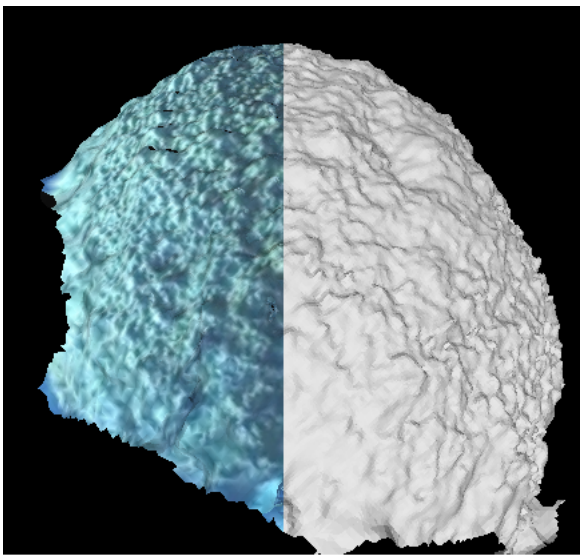
VII. Future Work

Having benchmarked our method with multiple control parameters, we are confident to deploy the technique into a more unstructured experimental environment. Extensions to the study presented here involve the analysis of *any* underwater video data sets collected in coral reef environments, as well as implementation of our algorithms onto underwater vehicles to facilitate automated data collection.

Initial steps have been taken to achieve the goal of utilising data from underwater video cameras to create 3-D reconstructions. In Fig. 11(a), we show a still-frame image of a coral head from a video sequence shot with a consumer video camera. Figure 11(b) shows the corresponding 3-D reconstruction from the subsequent footage captured by the camera. The left half of Fig. 11(b) is texture mapped with the coral's surface from the video footage, and the right half is a rendering of the underlying polygon mesh. Since the video was shot during a recreational dive, with no intent for reconstruction, the camera path was not ideal for image-based 3-D reconstruction. Although this reconstruction was done off-line, the techniques presented utilise no *a priori* knowledge of the camera or the survey environment, can learn the camera calibration on-the-fly, generate a dense mesh reconstruction of the feature, and can then compute the surface area and volume of a given coral colony or other reef structure. This initial experiment has provided motivation to continue pursuing the goal of automated and *in situ*, near-real time 3-D reconstruction of coral reef environments for applications in large-scale monitoring of growth and variability studies.



(a) Still frame image from a low-quality, consumer video camera of a coral head shot during a recreational dive.



(b) Three-dimensional reconstruction of the coral head shown in Fig. 11(a). The left half of the image is texture mapped with the coral's surface and the right half a rendering of the underlying polygon mesh object.

Fig. 11. Still frame image from a video sequence of a coral reef environment (Fig. 11(a)), with a corresponding 3-D reconstruction of the large coral head (Fig. 11(b)) taken from the video sequence.

Acknowledgment

This work was supported in part by the Early Career Academic Recruitment and Development (ECARD) program of the Queensland University of Technology. We would also like to thank Peter Mumby and Renata Legorreta for their direction, marine biology expertise, and the use of a physical coral specimen on which to conduct our experiments.

REFERENCES

[1] M. Spalding, C. Ravilious, and E. Green, *World Atlas of Coral Reefs*. University of California Press, Berkely, USA, 2001, prepared at the UNEP World Conservation Monitoring Centre.

[2] U.S. Environmental Protection Agency, "Coral reefs," http://water.epa.gov/type/oceb/habitat/factsheet.cfm, February 2011, viewed July 2011.

[3] D. Stoddart and R. Johannes, *Coral Reefs: Research Methods*. Paris: UNESCO, 1978.

[4] S. English, C. Wilkinson, and V. Baker, *Survey Manual for Tropical Marine Resources*, 2nd ed. Australian Institue of Marine Science, 1997.

[5] C. Rogers, G. Garrison, R. Grober, Z.-M. Hillis, and M. Franke, *Coral Reef Monitoring Manual for the Caribbean and Western Atlantic*. St. John: National Park Service, Virgin Islands National Park, 1994.

[6] J. Carleton and T. Done, "Quantitative video sampling of coral reef benthos: large-scale application," *Coral Reefs*, vol. 14, pp. 35 – 46, 1995.

[7] C. Laforsch, E. Christoph, C. Glaser, M. Naumann, C. Wild, and W. Niggl, "A precise and non-destructive method to calculate the surface area in living scleractinian corals using x-ray computed tomography and 3d modeling," *Coral Reefs*, vol. 27, pp. 811–820, 2008, 10.1007/s00338-008-0405-4. [Online]. Available: http://dx.doi.org/10.1007/s00338-008-0405-4

[8] A. Jones, N. Cantin, R. Berkelmans, B. Sinclair, and A. Negri, "A 3d modeling method to calculate the surface areas of coral branches," *Coral Reefs*, vol. 27, pp. 521–526, 2008, 10.1007/s00338-008-0354-y. [Online]. Available: http://dx.doi.org/10.1007/s00338-008-0354-y

[9] G. Holmes, "Estimating three-dimensional surface areas on coral reefs," *Journal of Experimental Marine Biology and Ecology*, vol. 365, no. 1, pp. 67 – 73, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022098108003729

[10] R. Vago, E. Vago, Y. Achituv, M. Ben-Zion, and Z. Dubinsky, "A nondestructive method for monitoring coral growth affected by anthropogenic and natural long term changes," *Bulletin of Marine Science*, vol. 55, pp. 126 – 132, 1994.

[11] K. Kruszyński, J. Kaandorp, and R. van Liere, "A computational method for quantifying morphological variation in scleractinian corals," *Coral Reefs*, vol. 26, pp. 831 – 884, 2007.

[12] O. Rahav, M. Ben-Zion, Y. Achituv, and Z. Dubinsky, "A photographic, computerized method for in situ growth measurements in reef-building cnidarians," *Coral Reefs*, vol. 9, p. 204, 1991.

[13] J. Bythell, P. Pan, and J. Lee, "Three-dimensional morphometric measurements of reef corals using underwater photogrammetry techniques," *Coral Reefs*, vol. 20, pp. 193 – 199, 2001.

[14] S. Cocito, S. Sgorbini, A. Peirano, and M. Valle, "3-d reconstruction of biological objects using underwater video technique and image processing," *Journal of Experimental Biological Ecology*, vol. 297, pp. 57 – 70, 2003.

[15] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, "Visual modeling with a hand-held camera," *International Journal of Computer Vision*, vol. 59, no. 3, pp. 207–232, 2004. [Online]. Available: http://www.springerlink.com/index/P5VK01158RK14252.pdf

[16] C. Olson, L. Matthies, H. Schoppers, and M. Maimone, "Robust stereo ego-motion for long distance navigation," vol. 2, 2000, pp. 453 –458 vol.2.

[17] P. Corke, D. Strelow, and S. Singh, "Omnidirectional visual odometry for a planetary rover," 2004.

[18] D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry," vol. 1, jun. 2004, pp. I–652 – I–659 Vol.1.

[19] K. Konolige, M. Agrawal, and J. Sola, "Large scale visual odometry for rough terrain," in *Proc. International Symposium on Robotics Research*. Citeseer, 2007. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.72.1240\&rep=rep1\&type=pdf

[20] K. Konolige and M. Agrawal, "Frameslam: From bundle adjustment to real-time visual mapping," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 1066–1077, 2008. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4648456

[21] C. Mei, G. Sibley, M. Cummins, P. Newman, and I. Reid, "A constant time efficient stereo slam system," in *Proceedings of the British Machine Vision Conference (BMVC)*, London, September 2009.

[22] M. Warren, D. McKinnon, H. He, and B. Upcroft, "Unaided stereo vision based pose estimation," in *Australasian Conference on Robotics and Automation*. ARAA, 2010.

[23] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1362–1376, 2010.

[24] V. Hiep, R. Keriven, P. Labatut, and J.-P. Pons, "Towards high-resolution large-scale multi-view stereo," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, june 2009, pp. 1430 –1437.

[25] N. D. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla, "Using multiple hypotheses to improve depth-maps for multi-view stereo," in *Proceedings of the 10th European Conference on Computer Vision: Part I*, ser. ECCV '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 766–779. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-88682-2_58

[26] C. Esteban and F. Schmitt, "Silhouette and stereo fusion for 3d object modeling," in *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, oct. 2003, pp. 46 – 53.

[27] C. Zach, "Fast and high quality fusion of depth maps," 2008.

[28] A. Zaharescu, E. Boyer, and R. P. Horaud, "Transformesh: a topology-adaptive mesh-based approach to surface evolution," in *In Proceedings of the Eighth Asian Conference on Computer Vision*, ser. LNCS 4844, vol. II. Tokyo, Japan: Springer, November 2007, pp. 166–175.

[29] P. Furgale and C. H. Tong, "Gpusurf - speeded up surf," Website, 2010, http://asrl.utias.utoronto.ca/code/gpusurf/index.html#sec-projects.

[30] B. Haralick, C. Lee, K. Ottenberg, and M. Nölle, "Review and analysis of solutions of the three point perspective pose estimation problem," *Computer Vision, International Journal of*, vol. 13, no. 3, pp. 331–356, 1994. [Online]. Available: http://www.springerlink.com/index/U450774M67121375.pdf

[31] G. Sibley, L. Matthies, and G. Sukhatme, "A sliding window filter for incremental slam," in *Unifying Perspectives in Computational and Robot Vision*, ser. Lecture Notes in Electrical Engineering, D. Kragic and V. Kyrki, Eds. Springer US, 2008, vol. 8, pp. 103–112.

[32] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 0, pp. 1–8, 2007.

[33] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02*, ser. CVPR '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 807–814. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2005.56

[34] N. Cornelis and L. V. Gool, "Real-time connectivity constrained depth map computation using programmable graphics hardware," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 1, pp. 1099–1104, 2005.

[35] P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, J.-M. Frahm, R. Yang, D. Nister, and M. Pollefeys, "Real-time visibility-based fusion of depth maps," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1 –8.

[36] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with," *Communications of the ACM*, vol. 24, no. 6, 1981.

[37] T. Boubekeur, P. Reuter, and C. Schlick, "Visualization of point-based surfaces with locally reconstructed subdivision surfaces," in *Proceedings of the International Conference on Shape Modeling and Applications 2005*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 23–32.

[38] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Eurographics Symposium on Geometry Processing*, 2006, pp. 61–70.

[39] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 519–528.

[40] C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," 2008.

[41] J. Bouguet, "Camera calibration toolbox for matlab," 2002. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/

[42] P. Besl and N. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on pattern analysis and machine intelligence*, pp. 239–256, 1992.

[43] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1 –8.